

## O ponto de inflexão em modelos de crescimento: Estimação, interpretação e aplicação a dados da COVID-19

Abraão da Silva Guimarães,<sup>1</sup> Valdemiro Piedade Vidas,<sup>2</sup> Erlandson Ferreira Saraiva<sup>3</sup>  
INMA, Campo Grande, MS

Devido ao cenário pandêmico causado pelo vírus SARS-CoV-2, causador da doença COVID-19, houve um aumento no interesse em modelagens estatísticas capazes de realizar projeções da evolução do número de casos (ou mortes) devido a COVID-19 em cidades/estados ou países. Esse interesse se deve ao fato de que as projeções podem auxiliar os governantes na tomada de decisões quanto à intensificação do isolamento social, aquisição de equipamentos hospitalares, aumento do número de unidades de terapia intensiva nos hospitais e/ou aplicação de um *lockdown*.

Especialmente no ano de 2020, muitos artigos foram publicados descrevendo procedimentos de modelagem para o número de casos e/ou mortes por COVID-19 em muitos países. Em geral, os trabalhos publicados modelam o número acumulado de casos (ou óbitos) por meio de algum modelo de crescimento não linear. Por exemplo, [1] considerou um modelo de crescimento exponencial para analisar a fase inicial da epidemia de COVID-19 na África; [2] considerou o modelo de crescimento logístico para o número acumulado de casos registrados na China e [3] considerou o modelo de Gompertz para prever o número de mortes no Brasil.

Neste texto, descrevemos uma modelagem estatística para dados de COVID-19 utilizando os modelos exponencial, Logístico e Gompertz. Primeiramente, deduzimos o modelo exponencial a partir de um exemplo simples, em que, uma quantidade de células dobra a cada unidade de tempo. A partir do modelo exponencial, escrito na forma de uma equação diferencial, e da suposição de um mecanismo limitador de crescimento, deduzimos os modelos Logístico e Gompertz. Deduzidas estas equações, obtemos o ponto de inflexão destes dois modelos. O interesse no ponto de inflexão é devido ao fato de que este representa o “pico” da pandemia para o período em estudo.

Assim, definindo  $N_t$  como sendo o número acumulado de casos até o instante de tempo  $t$ , temos que

$$N_t = \alpha_1 \exp\{\alpha_2 t\}$$

dá o número acumulado de casos em um crescimento do tipo exponencial;

$$N_t = \frac{\alpha_1}{1 + \alpha_2 \exp\{-\alpha_3 t\}}$$

dá o número acumulado de casos em um crescimento do tipo Logístico e

$$N_t = \alpha_1 \exp\{-\alpha_2 \exp\{-\alpha_3 t\}\}$$

dá o número acumulado de casos em um crescimento do tipo Gompertz, para  $t > 0$ .

Além disso, para evitar trabalhar com valores “altos” de  $N_t$ , optamos por considerar os modelos na escala logarítmica. Assim, tomando a transformação logarítmica em ambos os lados das três Equações e assumindo a presença de um erro aleatório, temos o seguinte modelo estatístico

$$Y_t = \log(N_t) = f(t) + \varepsilon_t,$$

<sup>1</sup>abraao.guimaraes@ufms.br

<sup>2</sup>valdemiro.vigas@ufms.br

<sup>3</sup>erlandson.saraiva@ufms.br

onde

$$f(t) = \begin{cases} \alpha_1 + \alpha_2 t & , \text{ se o modelo assumido é o exponencial;} \\ \alpha_1 - \log(\alpha_2 \exp\{-\alpha_3 t\}) & , \text{ se o modelo assumido é o Logístico;} \\ \alpha_1 - \alpha_2 \exp\{-\alpha_3 t\} d & , \text{ se o modelo assumido é o Gompertz;} \end{cases}$$

com  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ , para  $t > 0$ .

Para obter as estimativas para os parâmetros dos modelos, utilizamos o método dos mínimos quadrados não linear, conforme descrito por [4]. Este método está implementado no software R no pacote `nlstools`. Obtido as estimativas para os parâmetros dos três modelos, selecionamos o melhor modelo considerando como critério o critério de informação de Akaike, denotado por AIC, e o critério de informação bayesiana, denotado por BIC. O melhor modelo é aquele que possui menor valor AIC e BIC. Selecionado o melhor modelo, encontramos as coordenadas do ponto de inflexão e consequentemente a data estimada para o pico da pandemia no período considerado.

Ilustramos a modelagem descrita acima utilizando a quantidade de casos registrados na cidade de Campo Grande, MS, no período de 14 de Março de 2020 a 08 Novembro de 2020. Também aplicamos a modelagem ao número de mortes registradas no estado de São Paulo no período de 17 de Março 2020 a 30 de Abril de 2021.

No entanto, durante o processo de ajuste dos modelos, percebemos que os modelos considerados são adequados para modelar dados de apenas um período específico da pandemia, mas não todo o período considerado. Isto ocorre pois, em geral, o número de casos (ou mortes) por COVID-19 apresenta uma evolução heterogênea ao longo do tempo. Assim, como uma solução para este problema, consideramos o ajuste de um modelo por partes. Para identificar os diferentes períodos da pandemia ao longo do tempo, consideramos uma grade de valores com incrementos de tamanho 1, na qual, cada ponto da grade representa um possível ponto de mudança da pandemia. O incremento de uma unidade representa o acréscimo de um dia na pandemia. Em seguida, ajustamos os três modelos de crescimento para dois subconjuntos de dados obtidos de acordo com cada possível ponto de mudança. Selecionamos o melhor modelo para cada período de acordo com os valores AIC e BIC. O melhor ponto de mudança  $d$  é aquele associado ao modelo com o menor erro quadrático médio.

As duas principais vantagens em considerar o ajuste de um modelo por partes são: 1. sua eficácia na descrição dos diferentes períodos da pandemia; e 2. sua capacidade de explicar cada período da pandemia por meio das estimativas para os parâmetros epidemiológicos de interesse, como a taxa de crescimento do número de mortes diárias e a data de ocorrência do pico.

## Referências

- [1] S.S. Musa, S. Zhao e M.H. Wang. “Estimation of exponential growth rate and basic reproduction number of the coronavirus disease 2019 (COVID-19) in Africa”. Em: **Infect. Dis. Poverty** 9 (2020), p. 96. DOI: 10.1186/s40249-020-00718-y.
- [2] K. Wu e Q. Darcet D.; Wang. “Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world”. Em: **Nonlinear Dyn.** 101 (2020), pp. 1561–1581. DOI: 10.1007/s11071-020-05862-6.
- [3] J.A.M. Valle. “Predicting the number of total COVID-19 cases and deaths in Brazil by the Gompertz model”. Em: **Nonlinear Dyn.** 102 (2020), pp. 2951–2957. DOI: 10.1007/s11071-020-06056-w.
- [4] S. Vieira e R. Hoffmann. “Comparison of the Logistic and the Gompertz growth functions considering additive and multiplicative error terms”. Em: **Appl. Stat.** 26 (1977), pp. 143–148. DOI: 10.2307/2347021.